

DOCUMENT RESUME

ED 445 103

TM 031 785

AUTHOR Giraud, Gerald; Enders, Craig
TITLE The Effects of Repeated Cooperative Testing in an Introductory Statistics Course.
PUB DATE 2000-04-00
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Students; Cooperative Learning; Higher Education; *Statistics; *Student Attitudes; Study; *Test Results
IDENTIFIERS *Cooperative Test Taking

ABSTRACT

Cooperative testing seems a logical complement to cooperative learning, but it is counter to traditional testing procedures and is viewed by some as an opportunity for cheating and freeloading on the efforts of other test takers. This study examined the practice of cooperative testing in introductory statistics. Findings indicate that students had positive affect toward cooperative testing and also that they were not overly concerned with freeloading. Differences in self-reported study time between students who participated in cooperative testing and those who tested individually were not statistically significant. Students who participated in cooperative testing answered as many questions correctly on a non-cooperative final examination as did students in a traditionally tested course. While student performance, anxiety, and study time were not significantly affected by testing situation in this study, the study lends support to the use of cooperative testing in terms of student attitude. (Author/SLD)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ The document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to
improve reproduction quality.

• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

Gerald Giraud

1

The Effects Of Repeated Cooperative Testing In An Introductory Statistics Course

Gerald Giraud

Craig Enders

University of Nebraska - Lincoln

Paper presented at the Annual Meeting of American Educational Research Association, April,
1998, San Diego, CA

BEST COPY AVAILABLE

The Effects Of Repeated Cooperative Testing In An Introductory Statistics Course

Gerald Giraud, Ph. D. Nebraska Methodist College

Craig Enders, Ph. D. University of Miami, Florida

Key words: collaborative examinations; cooperative learning

Abstract: Cooperative testing seems a logical compliment to cooperative learning. However, cooperative testing is counter to traditional testing procedures, and is viewed by some as an opportunity for cheating and freeloading on the efforts of other test takers. This study examined the practice of cooperative testing in introductory statistics. Findings indicate that students had positive affect toward cooperative testing, and also that they were not overly concerned with freeloading. Differences in self reported study time between students who participated in cooperative testing were and those who tested individually were not statistically significant. Students who participated in cooperative testing answered as many questions correctly on a non cooperative final exam as did students in a traditionally tested course. While student performance, anxiety, and study time were not significantly affected by testing situation in this study, this study lends support to the use of cooperative testing in terms of student attitude.

Introduction

Cooperative testing, defined here as small group discussion of test items on the day of the exam, has been proposed as a logical extension of cooperative learning (Giraud, 1997). However, cooperative testing has not been advocated or studied to the same extent as cooperative learning. One potential reason for this is that cooperative testing is counter to traditional classroom testing practice, which often seeks to measure the learning of the individual for the purpose of assigning grades. In the minds of many teachers cooperative testing approximates what is traditionally viewed as cheating: Relying on others for answers. Another similar problem associated with cooperative testing is social loafing (Jackson & Harkins, 1985; Sears et al., 1988; Weber, 1992), also known as freeloading or slacking (Farland & Gullickson, 1984), which occurs when students take advantage of the efforts of other students.

Beside the potential negative aspects of cooperative testing are potential benefits. Cooperative testing is a logical compliment to cooperative learning. The high stakes nature of the testing situation might increase the benefits of cooperation on the learning process by enhancing the interaction between learners. Because students have the opportunity to discuss the test items with members of a familiar cooperative group, anxiety and test day discomfort might also be reduced.

In sum, the high stakes nature of a test should encourage cooperation and facilitate learning. However, cooperative testing might encourage students to depend on others in the group for answers, potentially resulting in reduced study of course material -- a practice that might interfere with learning. If cooperative testing is to be considered as a component of cooperative learning, it is important to consider whether students learn and retain material when tested in cooperative groups. If students retain as much or more of course material when they are tested cooperatively as they do when tested individually, then cooperative testing practices are supported. However, cooperative testing cannot be justified if students cannot retain course material when testing cooperatively.

Cooperative Testing in College Classrooms

Only a few studies have examined cooperative testing in the college classroom. Farland and Gullickson (1984) studied cooperative testing in a measurement class for college seniors. Forty-six students were randomly assigned to two groups, and then randomly assigned to cooperative learning groups of 4 or 5. The authors acknowledged the possibility of freeloading, and hoped that random assignment, along with individual general examinations, would help control it. Students worked on projects in the small groups. One section took frequent five-item cooperative quizzes, while students in the other section were quizzed individually. Students in both sections completed two general examinations individually. While Farland and Gullickson found that students liked cooperative tests and felt that the tests helped them learn, they found no difference in the scores of the two groups on the general examinations, and little difference between quiz scores. Cooperation was not tried in the general examinations. Neither study time nor perceptions of freeloading were examined.

Meinster and Rose (1993) studied cooperative testing with two classes of introductory

psychology students. Students could form pairs and test cooperatively for two of the four multiple choice classroom tests; students were also free to test individually. Students who chose to form pairs reported positive affect toward cooperative testing. This study did not involve long term cooperative learning groups that cooperated in testing, and did not address the problem of freeloading.

Giraud (1997) studied cooperative testing in both a graduate and an undergraduate introductory statistics course. Undergraduate students were assigned to cooperative learning groups. During one test of the semester, students were given the opportunity to discuss test items before completing the test individually. Graduate students were given a take-home test and encouraged to collaborate. Time was also allotted during class for test discussion. In general, Giraud found that students reported studying the same amount of time or longer for cooperative tests as they did for individually completed tests. Undergraduates acknowledged the possibility for freeloading, but responded positively to cooperative testing nonetheless. Graduate students reported feeling obligated to prepare more thoroughly because of the cooperative component: They wanted to contribute to the discussion and appear knowledgeable before peers. This study examined responses to cooperative testing for only a single testing occasion.

Previous studies have not examined cooperative testing or its affect on study time, student attitudes and learning across repeated occasions. The current study seeks to add to understanding of cooperative testing by addressing the following questions:

1. Do student attitudes toward cooperative testing change after repeated exposures to the testing format?
2. Does the amount of self-reported study time change after repeated exposures to cooperative testing?
3. What are the students' perceptions of freeloading in the collaborative testing environment?
4. Do perceptions of freeloading change after repeated exposures to cooperative testing?
5. Do students tested individually differ in terms of test anxiety from students who are tested cooperatively?
6. Do students tested individually differ in terms of retention of course material from students who are tested cooperatively?

Method

Participants

Participants in this study were 53 undergraduate students enrolled in one of two consecutive summer sessions of an undergraduate introductory social science statistical methods course at a large Midwestern university. As the course fulfills university-wide requirements, students came from diverse academic backgrounds. At the time of enrollment students were unaware of the study.

Treatment

A traditional individual testing format was used in the first session class, while cooperative testing was used in the second summer session class. The assignment of the

treatment conditions was determined randomly. Instruction was provided by the same instructor during both terms, and differed only with respect to the testing method.

At the beginning of the course, students in both classes were randomly assigned to cooperative learning groups of three or four students. These groups remained the same throughout the course. Daily class meetings for both sections involved both lecture and group learning components: The first half of class meeting time was devoted to lecture, and the remainder of class was spent on cooperative learning activities related to the lecture.

Three tests were administered to both classes. The first two tests covered only new material. The final test also covered new material, but included an additional comprehensive section comprised of 13 multiple choice items. The three exams were identical for both classes. Prior to taking each test, all students completed a brief questionnaire designed to measure test anxiety (see appendix). Students were also asked to report how much time that they spent preparing for the exam.

Prior to each exam, students in the cooperative testing section were given the opportunity to discuss item stems in their assigned cooperative learning groups. Students were given test forms with item stems, but not answer choices, and were allowed 15 minutes to discuss the test items. When the 15 minutes had expired, students were given a complete test form (stems and response choices) that was completed individually. Following each test, students in the cooperative testing class completed a brief questionnaire designed to measure their attitudes toward the cooperative testing experience. In addition, students were also asked to compare their preparation time to time spent on similar, non-cooperative tests (see appendix). Because the retention of course content, as measured by the comprehensive section of the final test, was of interest in this study, it is important to note that all students completed the 13 question comprehensive section of the final without the benefit of group discussion.

Analysis

Change in student attitudes toward cooperative testing across the three test administrations was assessed using a repeated measures ANOVA. The dependent variable for this analysis was a scale score created by averaging responses across the items of the cooperatively testing questionnaire. As the questionnaire items were measured using a five-point Likert scale, the resulting scale score also had a range of five points. The change in self-reported study time was assessed using a split-plot ANOVA (2 conditions by 3 test occasions). It is reasonable to assume that study time might change differentially for the cooperative and traditional conditions. Therefore, it was of particular interest to examine the interaction term for this analysis. Student perceptions of freeloading behavior were examined using descriptive statistics from the third and fourth questionnaire items. The change in the perceptions of freeloading across testing administrations was assessed using a Friedman two-way analysis of variance by ranks, a non-parametric test analogous to repeated measures ANOVA that is appropriate for use with ordinal data. The effects on test anxiety were assessed using a split-plot ANOVA (2 conditions by 3 test occasions). Because the level of test anxiety might differentially change across the testing occasions for the two groups, it was again of interest to examine the interaction term. The dependent variable for this analysis was a scale score created by averaging responses across the items of the test anxiety questionnaire. The resulting scale score had a five-point range, where higher values denoted more anxiety. Finally, retention of course content was

assessed using a t-test to compare the mean percentage correct on the cumulative section of the final test for individual and cooperative test classes. It should also be noted that an effect size measure, eta squared (η^2), was examined for all parametric analyses. Eta squared can be interpreted similar to r^2 in a correlational setting. Cohen (1977) suggested that η^2 values of .01, .06, and .14 could be interpreted as small, medium, and large effect sizes, respectively.

Results

Attitudes Toward Cooperative Testing

The proportion of 'agree' or 'strongly agree' responses for each item of the attitude scale are reported in Table 1. Means and standard deviations for the attitude toward cooperative learning scale across the three test administrations are reported in Table 2. The internal consistency reliability of this instrument was estimated at each of the three administrations. The average coefficient alpha was .73. A repeated measures ANOVA revealed a statistically significant increase in affect toward cooperative testing across the three exams ($F = 3.68$, $p = .039$). Based on Cohen's (1977) benchmark values, this change in attitudes represented a large effect size ($\eta^2 = .22$). Follow-up tests indicated a statistically significant increase between both exam 1 and 2 as well as exam 2 and exam 3.

Changes in Study Time

Students' report of study time for the cooperative tests relative to similar, non-cooperative tests is reported in Table 3. Means and standard deviations for self-reported study time are reported in Table 4. A split-plot ANOVA was used to compare the change in study time for the two groups. Results indicated that both the interaction and the between-subjects main effect, testing condition, were not statistically significant. The effects sizes for the interaction and testing condition main effect were both small: $\eta^2 = .054$ and $\eta^2 = .019$, respectively. As seen in Table 4, self-reported study time does appear to be larger for traditional testing group. However, the large variability in study time makes these differences non-significant. As seen in the table, study time did appear to decrease across the three test administrations for the cooperative testing section. It should be noted that the within-subjects main effect was statistically significant ($\eta^2 = .104$). Because this effect was not of particular interest to the study it will be ignored.

Perceptions of Freeloading

Perceptions of freeloading were measured by questions 3 and 4 on the attitudes toward cooperative testing questionnaire (see Table 1). Because freeloading was measured using two ordinal items, the change in perceptions across testing administrations for the cooperative testing group was assessed using a Friedman two-way analysis of variance by ranks, a non-parametric test analogous to a repeated measures ANOVA. Results indicated no statistically significant change in freeloading perceptions across the three exams. Although not statistically significant, it should be noted that a consistent trend was observed for both items; perceptions of freeloading increased across the three test administrations.

Test Anxiety

Means and standard deviations for the test anxiety scale are reported in Table 5. The internal consistency reliability of this instrument was estimated at each of the three

administrations. The average coefficient alpha was .94. A 2 x 3 split-plot ANOVA was used to compare the changes in test anxiety for the two groups. Results indicated that none of the effects in the design were statistically significant. Furthermore, the effect sizes for the interaction term and the within-subjects main effect were negligible: $\eta^2 = .001$ and $\eta^2 = .007$, respectively. The effect size for the between-subjects main effect was small ($\eta^2 = .041$). Although not statistically significant, the cooperative testing group experienced somewhat lower levels of test anxiety than the traditional testing group did (2.91 versus 3.123, respectively).

Retention of Course Material

Means and standard deviations for the 13 item cumulative final exam are reported in Table 6. An independent t-test revealed no statistically significant differences between the two testing groups ($\eta^2 = .034$). Although not statistically significant, the cooperative testing section scored somewhat lower on the cumulative portion of the final than the traditional testing group did. Although the mean percentages shown in Table 6 appear to indicate a fairly substantial difference between the two groups, it should be noted that this difference represents only half an exam point.

Discussion

The current study investigated the effects of repeated cooperative testing on five dependent variables: attitudes toward cooperative testing, self-reported study time, perceptions of freeloading, test anxiety, and retention of course material. The following results were obtained:

1. Students' attitudes toward cooperative testing became more positive after each test administration.
2. Self-reported study time varied substantially among students. The mean study time was less (but not statistically significant) for the cooperative testing group. A minority of students reported studying less for the cooperative tests as compared to similar, traditional exams.
3. Students' perceptions of freeloading increased across test administrations. This trend was not statistically significant.
4. The cooperative testing sections appeared to experience slightly less test anxiety than did the traditional testing section. Again, this finding was not statistically significant.
5. Testing condition did not appear to affect retention of course material. Although the cooperative testing group scored slightly lower on the cumulative final, the difference was not statistically significant.

The current study lends support to the use of cooperative testing on the grounds of student preferences; students' attitudes toward cooperative testing became more positive after each successive exposure. Another important benefit of using cooperative testing might be a reduction in test anxiety. Although this effect appears to be small, even minor reductions in anxiety might be substantively important. For an undergraduate population, one of the potential drawbacks of cooperative testing is freeloading behavior. Although the effect was not dramatic, students did perceive freeloading to be more problematic after repeated exposures to cooperative testing. However, if freeloading artificially inflated some students' performance, it might follow

that test scores on the cumulative final would have been lower for cooperatively testing section, as they were not allowed to discuss these items. This did not appear to be the case. Furthermore, if less prepared students were obtaining answers from the more prepared students, we might expect that overall test scores would have been higher for the cooperative testing section. This was not the case. With respect to study time and retention, results were somewhat inconclusive. In general, the effects of cooperative testing studied herein appear to be small. Although the effect sizes were generally small, it does not necessarily follow that differences between the two groups are substantively unimportant.

Limitations

This study suffered from several limitations. Perhaps the most significant limitation of this study was low power. As previously noted, the initial sample was comprised of only 53 students. Unfortunately, non-response on some of the dependent measures resulted in a further reduction of the sample size, and subsequently statistical power. The lack of statistical power was further complicated by small effect sizes. As noted in the previous section, most of the effect sizes found in this study were small. A dramatically larger sample would have been required to obtain statistically significant results in this situation.

A second potential limitation of this study was insensitive measurement devices. Specifically, the 13 item cumulative test used to measure course retention appeared to suffer from a ceiling effect; the exam was too easy. A longer, more difficult cumulative test might have been more sensitive to between-group differences. The fact that both sections were taught during a five-week summer session might have contributed further to this problem. In addition, the use of self-reported study time as a dependent variable resulted in large within-groups variability. In this case, the use of a study log might have resulted in a reduction in error variance.

A general caveat should also be noted. Although the treatment condition was randomly assigned to one of the two sections, students determined which section they enrolled in. Although students were not aware of the study at the time of enrollment, it is still possible that the two sections natural differed on some important, unmeasured dimension. Viewed this way, results should be generalized with some caution.

Future Research

Participants in the current study were undergraduate students. A previous study by Giraud (1997) suggested that graduate students might benefit more from cooperative testing than undergraduate students would. In Giraud's study, graduate students reported feeling more obligated to prepare for the cooperative exam. As a result, the mechanisms that influence freeloading might be completely different for graduate students as opposed to undergraduates. As a result, extending this study to a population of graduate student would be of considerable interest. Also with respect to freeloading, it might be of interest to change the cooperative groups prior to each exam in future studies. It is possible that freeloading might have increased as a result of maintaining the same cooperative learning groups through the entire semester; students may have felt increasingly comfortable depending on their group members as time passed. Also, future studies should attempt to address the power and measurement problems discussed above. If more sensitive instruments could be implemented, more conclusive results might be obtained. Finally, future studies might also examine whether or not differences in freeloading occur between traditional lecture classes and cooperative classes when cooperative

testing is used. If students in lecture classes freeload more than students in cooperative classes, it would suggest that cooperative learning promotes pro-social behavior. The effects of cooperative testing on group interactions, affect toward group members, and conceptual understanding might also be interesting topics for further study.

References

- Cohen, J. (1977). Statistical power for the behavioral sciences. New York: Academic Press.
- Farland, D. & Gullickson, A. (1984) Cooperative test-taking. Paper presented at the annual meeting of the Northern Rocky Mountain Research Association, Jackson, WY.
- Giraud, G. (1997) Cooperative testing in a cooperative learning classroom. Paper presented at the annual meeting of the American Educational Research Association, Chicago, ILL.
- Jackson, J.M. & Harkins, S.G. (1985) Equity in effort: An exploration of the social loafing effect. Journal of personality and social psychology, 49 (5), 1199-1206.
- Meinster, M. & Rose, K. (1993) Cooperative testing in introductory level psychology courses. Proceedings of the annual conference on undergraduate teaching of psychology, Ellenville, NY.
- Sears, D.O; Peplau, L.A.; Freedman, J.L.; and Taylor, S.E. (1988) Social psychology, (pp. 385-386) Englewood Cliffs, NJ: Prentice Hall.
- Weber, A.L. Social Psychology, (pg. 181) New York: HarperCollins.